DOCUMENT RESUME

ED 246 058 TM 840 239

AUTHOR Carroll, C. Dennis
TITLE The Effect of Alternative Definitions of Growth on Group Differences in Achievement.
PUB DATE 24 Apr 84
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Gains; Achievement Tests; *Pretests Posttests; Program Effectiveness; *Research Problems; Scores; Secondary Education; *Statistical Analysis
IDENTIFIERS *Gain Scores; High School and Beyond (NCES)

ABSTRACT

This paper compares five growth measures using the High School and Beyond database. The measures are: (1) simple gain (posttest minus pretest); (2) difference between group means; (3) percentage of students who scored higher on the posttest than the pretest; (4) percentage of items missed on the pretest which are subsequently answered correctly on the posttest; and (5) difference between the observed posttest scores and the predicted posttest scores based on the correlation between the pre- and posttest scores. These methods were evaluated by determining how consistent each is in differentiating between groups which, on the basis of other evidence, are almost certain to differ in a specified direction. The groups considered were based on sex, race/ethnicity, socioeconomic status, and ability level. Problems were found with all five measures of growth. It is recommended that researchers use multiple measures to study growth. If that is not possible, the percentage initially wrong but subsequently correct measure or the regression residual should be used. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED246058

SESSION 16.05
PROBLEMS IN ASSESSING COGNITIVE GROWTH IN HIGH SCHOOL:
EVIDENCE FROM HIGH SCHOOL AND BEYOND

THE EFFECT OF ALTERNATIVE DEFINITIONS OF GROWTH ON GROUP DIFFERENCES IN ACHIEVEMENT

Dr. C. Dennis Carroll, Chief
Longitudinal Studies Branch
National Center for Education Statistics

TM 840 239

Paper Presented to the 1984 Annual Meeting of the
American Educational Research Association
April 24, 1984

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Sylvia Butler

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

PROBLEMS IN ASSESSING COGNITIVE GROWTH IN HIGH SCHOOL: EVIDENCE FROM HIGH SCHOOL AND BEYOND

THE EFFECT OF ALTERNATIVE DEFFINITIONS OF GROWTH ON GROUP DIFFERENCES IN ACHIEVEMENT

Dr. C. Dennis Carroll, Chief
Longitudinal Studies Branch
National Center for Education Statistics

Growth has frequently presented problems to researchers in education, psychology, sociology, and other sciences concerned with learning. Many of these problems were identified several decades ago, but the solutions are difficult to find. In general, the problems fall into categories which can be loosely labeled as: reversal, attenuation, dimensionality, and group bias. Reversal refers to shifts in growth measures as an inverse function of initial scores. Attenuation refers to a diminution or shrinkage in observed growth which arises from errors of measurement in both the pre- and posttests that error-free measurements would have revealed. Dimensionality problems are found when the pre- and post-measures have different meanings--as is frequently the case when a lengthy time period separates observations. Finally, group biases stem from differing interpretations of items by subgroups of learners. Many researchers doubt that these problems will ever be solved.

In spite of the problems, growth is probably the most important construct in the field of learning. Educational policy-makers at all levels are frequently faced with decisions that benefit from estimates of growth. When these decisions are based on a single growth measure, many observers (and frequently the decision-maker) feel uncomfortable. This paper will only inform or

illuminate these feelings of discomfort but is not likely to ease the discomfort.

The High School and Beyond data base offers a unique opportunity to study growth. It is derived from a large, nationally representative sample of students and schools. High School and Beyond included tests specifically designed to facilitate growth measurements and the pre- and post-observations were timed in such a fashion that most people would expect some growth to have occurred.

This paper will compare several growth measures using the High School and Beyond data. No effort is made to consider all possible measures of growth, or to identify all of the problems. The main concern of this paper is to evaluate various methods of measuring growth by determining how consistent each is in differentiating between groups which, on the basis of other evidence, are almost certain to differ in a specified direction. To this end, we will compare five different tests, and look at the effects which each measure has on certain group differences. The groups to be considered are based on (a) sex, (b) race/ethnicity, (c) socioeconomic status (SES), and (d) ability level.

High School and Beyond Data

The High School and Beyond study began in spring 1980 with about 28,000 seniors and 30,000 sophomores enrolled in 1,015 public and private high schools across the nation. The sophomore cohort provided the data for the analyses in this paper. The first follow-up of these sophomores took place in spring 1982, when most were seniors. Every sophomore cohort sample member

still in the same school was retained with certainty; all others were subsampled. The resulting first follow-up survey sample size was about 30,000.

Questionnaires and cognitive tests were administered to each student in the sample in the base-year survey and again in the first follow-up survey. The student questionnaire covered school experiences, activities, attitudes, plans, background characteristics, and language proficiency. The sex, race/ethnicity, and socioeconomic status variables employed in this paper were derived from the questionnaire responses. The cognitive tests--the same ones were administered in both the base-year and first follow-up--were developed by the Educational Testing Service of Princeton, New Jersey especially for the sophomore cohort of High School and Beyond. A committee of experts judged the items to be sensitive to growth that is likely to occur in the last two years of high school. The tests included vocabulary, reading, writing, computational mathematics, algebra and geometry, science, and civics. The ability levels incorporated in these analyses were quartiles of the average standardized scores on the base-year reading, mathematics, and vocabulary tests.

Analyses were restricted to the 22,436 sophomores responding to both tests (base-year and first follow-up) and the needed questionnaire items.

The cognitive tests used in this study were the computational mathematics and the science test. The computational mathematics test consists of 28 items administered with a 16-minute time limit. Each item requires a judgement of equality or direction of inequality and basic algebraic computations. The concepts covered

by the mathematics test include simple equations in one unknown, value substitution, exponentiation, coordinate reading, linear measurement, and area and volume calculations. Only four of the items are worded with algebraic symbols. The science test consists of 20 items administered with a 10-minute time limit. Each item requires the selection of the best answer from five choices. Eight of the items are related to biology and the remaining twelve items are related to physical sciences (two are related to astronomy, five are related to chemistry, two are related to geology, and two are related to physics).

Growth Measures

This paper examines five measures of growth. The simplest of all the measures is the simple gain--posttest minus pretest. This measure has two virtues that make it intuitively appealing: (1) it capitalizes on the longitudinal character of the data, and (2) it benefits from the use of within groups rather than between groups variance. However, this measure is plagued by so many problems that few researchers dare use it. Instead, they elect to use difference between group means as their measure of growth. This measure requires an analysis of interactions for subgroup differences, which is sometimes difficult to explain. The third measure to be considered is very crude. In fact, it can only indicate whether growth occurred, but not its magnitude. I'm referring to the measure obtained by calculating the percentage of subjects who scored higher on the posttest than on the pretest. This measure is far from perfect. Indeed, it has an expected value of 50 in the presence of zero growth.

The fourth measure requires item computations. It is the percentage of all items missed on the pretest which are subsequently answered correctly on the post-test. The intuitive appeal of this measure should be balanced against the difficulty of the calculations required and the complexity of the definition, especially for laymen.

The fifth and final measure considered in this paper is a regression residual, that is, the difference between the observed posttest scores and the predicted posttest scores based on the correlation between the pre- and post-test scores. Problems with this measure stem from the estimation of the relationship (correlation) between the pre- and posttest scores. Most estimates of the correlation are biased, and most corrections for the bias only compound the problems. In addition, this measure is difficult to explain to nontechnical audiences.

Ceiling Problems

All of these growth measures behave poorly when the students perform very well, on either the pre- or the posttest. Simple gain scores fail to show growth and even show false declines when the pretest scores approach the ceiling. Group mean differences lose power due to restrictions in variance. The percentage showing any growth becomes zero when at the ceiling (in spite of an expected value of 50 percent). The demoninator of the percentage of the initially wrong items subsequently answered correctly measure is zero at the ceiling. Finally, the regression estimates become very dubious due to restriction of range near the ceiling.

Table 1.--Cumulative percentage distributions for the High School and Beyond computational mathematics and science tests, base-year and first follow-up

Test value	Unweighted		Weighted	
	Base-year	First follow-up	Base-year	First follow-up
Computational Mathematics				
Missing	0.5	1.8	0.4	1.7
14	65.0	56.8	64.7	56.8
16	75.0	66.0	74.9	65.9
18	80.3	71.1	80.4	71.3
20	88.0	80.1	88.1	80.4
22	91.8	84.9	91.9	85.3
24	96.6	92.1	96.7	92.2
25	96.7	92.3	96.9	92.4
26	98.4	95.6	98.6	95.6
27	99.5	98.2	99.6	98.2
28	100.0	100.0	100.0	100.0
Science				
Missing	1.3	2.9	1.2	2.8
10	59.1	51.0	57.8	50.0
12	70.8	62.3	69.9	61.3
14	86.9	80.5	86.5	79.8
16	93.0	88.9	92.9	88.6
17	96.9	94.4	96.8	94.3
18	98.9	98.0	98.9	97.9
19	99.8	99.6	99.7	99.7
20	100.0	100.0	100.0	100.0

The computational mathematics test and the science test used in High School and Beyond do not appear to suffer from ceiling problems. As Table 1 shows, less than one percent of the students had perfect scores on these tests in the base-year. Furthermore, less than four percent of the students were within four items of the ceiling (scores of 25 or more) on the base-year mathematics test. Less than seven percent of the students were within four items of the ceiling on the base-year science test.

The growth measures used in this paper were adjusted for the minor ceiling effects problem by deleting from all analyses those students with perfect pretest scores. After all problematic cases (ceiling and missing item data) were removed, the analyses discussed in this paper were based on 20,309 observations.

RESULTS

The five growth measures were calculated for four groupings of students. The groupings were on the basis of ability, socioeconomic status, race/ethnicity, and sex. Differences, and patterns of differences, were evaluated by means of t-tests which were calculated through the use of balanced repeated replicate standard errors. These results are presented below for each grouping.

Ability

Table 2 displays the results for the ability subgroups. Ability subgroups were defined by the quartiles of the base-year composite test scores. The "middle" grouping includes the middle two quartiles.

As would be expected, the low ability students scored substantially lower than the middle ability students who in turn scored substantially lower than the high ability students on both the pre- and the post-tests. The large initial differences on both the computational mathematics test and the science test are comforting (in the sense that smarter students performed better). The tests are at least discriminating between groups of students that theoretically are different.

Table 2.--Growth in computational mathematics and science, by ability grouping

Index	Mathematics			Science		
	Low	Middle	High	Low	Middle	High
Base-year formula score						
Mean	1.71	9.60	19.00	4.24	8.94	13.26
s.e.	.07	.08	.08	.06	.04	.05
t a/	74.22	162.65	83.08	65.17	115.48	67.46
First follow-up formula score						
Mean	3.46	11.14	20.44	5.38	9.92	14.03
s.e.	.09	.10	.11	.05	.05	.05
t	57.08	119.47	62.55	64.20	122.32	58.12
Simple gain scores						
Mean	1.75	1.54	1.44	1.14	.98	.77
s.e.	.09	.07	.11	.06	.05	.04
t	-1.84	-2.18	-0.76	-2.04	-5.13	-3.27
Group mean differences						
Mean	1.75	1.54	1.44	1.14	.98	.77
s.e.	.11	.12	.13	.08	.06	.07
t	-1.22	-1.75	-0.54	-1.58	-3.51	-2.20
Percentage scoring higher on the posttest						
Mean	61.79	60.92	67.10	57.92	57.29	56.13
s.e.	.95	.70	.98	.84	.59	.75
t	-0.74	3.89	5.13	-0.61	-1.59	-1.22
Percentage initially wrong, but subsequently correct						
Mean	26.41	39.39	59.08	29.53	37.26	45.02
s.e.	.23	.30	.46	.25	.32	.35
t	34.33	63.52	35.85	19.04	36.01	16.36
Regression residuals						
Mean	-0.90	-0.21	0.74	-1.21	-0.12	0.81
s.e.	.08	.06	.11	.06	.05	.04
t	6.90	12.05	7.58	13.96	28.01	14.52

a/ The t-values placed in the "Low" columns are for comparisons of middle quartile students with low quartile students. The t-values in the "Middle" columns are for comparisons of high and low quartile students. The t-values in the "High" columns are for comparisons of high and middle quartile students.

With respect to growth, however, the picture is less clear. The mean gain scores earned by the low, middle, and high ability groups differed significantly from each other for the science test but not consistently so for the mathematics tests. It is disturbing to note, however, that with respect to both tests, the

mean gain scores earned by the different ability groups are inversely related to pretest score (the least able students gained the most) 1/. In other words, the phenomenon of reversal plagues both tests.

The absolute size of the simple gains is relatively small for both tests. For mathematics, 1.75 out of 28 items is not very much growth, considering that two years elapsed between the pre- and post-tests. The 1.14 out of 20 items gain on the science test is even more discouraging. These absolute sizes are comforting with respect to problems of unidimensionality. That is, it is doubtful that the underlying content of the tests changed given the highly similar pre- and post-test scores. However, the problem of attenuation may be very serious for both the mathematics and the science tests. The simple gains certainly appear to be attenuated.

As expected, the difference in group means as a growth measure yielded results quite similar to those produced by simple gain scores. In fact, algebraically, these two measures are, of course the same. As such they both suffer from the reversal problem mentioned above and probably from the attenuation problem as well. The use of between errors, in the group means measure as contrasted with within errors as used for the simple gains, resulted in larger standard errors and lower t-values. For this reason, the group means measure suppresses two significant differences that had shown up with the simple gains measure.

1/ Psychometricians expect reversal for many measures of growth (e.g., weight gain or loss)--but, most educators are disturbed by this in learning situations.

The third measure, percentage scoring higher in the posttest, displays an entirely different pattern from that shown by the first two measures. The raw percentages appear reasonably high, until the expected value is considered. If the theoretical expected value of about 50 percent is subtracted from each percentage, then relatively low values are obtained. Certainly, the reversal problem is evident for the science test, and an inconsistent pattern is evident for the mathematics test. It is likely that this measure suffers from both attenuation and reversal. The pattern of t-values is quite different for this measure. For the science test none of the comparisons is significant, but for the mathematics test two of the three are significant. This pattern does not agree with the simple gain measure for any of the comparisons. Indeed, in the one instance where both measures identify a significant difference (between high and low quartile students on the mathematics test) the direction of the difference is different for the two measures. It is only this third measure (percentage scoring higher on the posttest) that yeilds a difference in the expected direction, i. e. with the high ability group scoring higher.

The fourth measure is percentage of items missed on the pretest but answered correctly on the posttest. This measure appeared to work very well. Reversal was not a problem and the magnitude of the values was acceptable enough to discount attenuation problems. All six t-values are clearly significant and in the theoretically expected directions.

The regression residuals appeared to work reasonably well. No reversal problems were identified, but the small values suggest

that attenuation may have been a problem. However, all six t-values are significant and in the theoretically expected directions but are not as large as those calculated for the previous measure.

Overall, the best measure appears to be the fourth one: percentage of items initially wrong, but subsequently correct although the regression residuals measure seemed to work reasonably well. The other three measures were relatively problematic, with the group mean differences measure displaying the most problems.

Socioeconomic Status

Socioeconomic status is a well known classification variable. It has consistently shown relationships to most measures of ability, learning, and other cognitive indices. It was included in this paper to represent possible independent variables.

As Table 3 shows, socioeconomic status was indeed related to performance on both mathematics and science in both the pre- and posttest administrations. The simple gain measure and the percentage scoring higher on the posttest indicated significant growth differences for the mathematics test but not for the science test. The group mean difference measure was substantially weaker, as discussed above. However, the percentage initially wrong but subsequently correct and the regression residual measures indicated significant growth differences for both the mathematics and science tests.

It is interesting to note that the relative power of two measures (regression residuals and the percentage initially wrong but subsequently correct) was opposite for the two tests. The

Table 3.--Growth in computational mathematics and science by socio-economic status

Index	Mathematics			Science		
	Low	Middle	High	Low	Middle	High
Base-year formula score						
Mean	6.96	10.59	14.35	7.05	9.31	11.17
s.e.	.17	.11	.18	.10	.06	.12
t a/	17.93	29.84	17.82	19.38	26.37	13.86
First follow-up formula score						
Mean	7.98	12.12	16.50	7.95	10.30	12.13
s.e.	.15	.13	.22	.11	.06	.10
t	20.86	31.99	17.14	18.75	28.11	15.69
Simple gain scores						
Mean	1.02	1.53	2.15	.90	.99	.96
s.e.	.12	.08	.09	.05	.05	.05
t	3.54	7.53	5.15	1.27	0.85	-0.42
Group mean differences						
Mean	1.02	1.53	2.15	.90	.99	.96
s.e.	.23	.17	.28	.15	.08	.16
t	1.80	3.11	1.87	0.53	0.28	-0.17
Percentage scoring higher on the posttest						
Mean	57.48	61.99	69.42	56.68	57.42	57.12
s.e.	1.27	.65	.91	.67	.72	.75
t	3.16	7.64	6.64	0.75	0.43	-0.28
Percentage initially wrong, but subsequently correct						
Mean	33.75	41.90	51.34	33.52	38.39	41.41
s.e.	.38	.36	.65	.40	.29	.34
t	15.57	23.36	12.70	9.86	15.03	6.76
Regression residuals						
Mean	-1.04	-0.13	0.90	-0.72	-0.01	0.44
s.e.	.11	.08	.10	.05	.05	.04
t	6.69	13.05	8.04	10.04	18.12	7.03

a/ The t-values placed in the "Low" columns are for comparisons of middle quartile students with low quartile students. The t-values in the "Middle" columns are for comparisons of the high and low quartile students. The t-values in the "High" columns are for comparisons of high and middle quartile students.

percentage initially wrong but subsequently correct measure yielded larger t-values than the regression residuals for the mathematics test, but the opposite was found for the science test.

Race/ethnicity

Table 4 displays the pattern of differences for whites, blacks, and Hispanics. The whites scored higher than the blacks and Hispanics on both the mathematics and the science test for both the pre- and post-test administrations. However, the growth measures showed a pattern distinctly different from any seen in the previous tables.

Overall, the group mean difference measure failed to identify differences as would be expected. The simple gains measure and the percentage with any growth measure each produced a very different pattern from that produced by the regression residuals measure and percentage initially wrong but subsequently correct measure. Indeed, the simple gains and percentage scoring higher on the posttest measures failed to reveal any differences between whites and blacks, but the regression residuals and percentage initially wrong but subsequently correct measures identified large significant differences. On the other hand, the simple gains and the percentage scoring higher on the post-test measures for the mathematics test identified significant differences between the blacks and Hispanics, but the percentage initially wrong but subsequently correct measure did not. The regression residuals measure identified a significant difference, but the t-value was smaller. For the science test, the regression residuals and the percentage initially wrong but subsquently correct measures found significant differences between the blacks and Hispanics, but not the simple gains or the percentage scoring higher on the posttest measure.

Table 4.--Growth in computational mathematics and science by race/ethnicity

Index	Mathematics			Science		
	White	Black	Hispanic	White	Black	Hispanic
Base-year formula score						
Mean	12.20	5.36	6.22	10.23	5.53	6.47
s.e.	.12	.20	.19	.07	.15	.10
t a/	-29.33	-26.61	3.12	-28.39	-30.80	5.21
First follow-up formula score						
Mean	13.87	6.84	7.20	11.21	6.40	7.35
s.e.	.14	.24	.22	.06	.11	.11
t	-25.30	-25.58	1.11	-38.39	-30.81	6.11
Simple gain scores						
Mean	1.67	1.48	.98	0.98	0.87	0.88
s.e.	.07	.11	.12	.04	.10	.08
t	-1.46	-4.97	-3.07	-1.02	-1.12	.08
Group mean differences						
Mean	1.67	1.48	.98	0.98	0.87	0.88
s.e.	.18	.31	.29	.09	.19	.15
t	-0.52	-2.00	-1.17	-0.53	-0.57	0.04
Percentage scoring higher on the posttest						
Mean	63.90	61.81	56.71	57.11	56.36	56.74
s.e.	.62	1.17	1.20	.47	1.22	1.06
t	-1.58	-5.32	-3.04	-0.57	-0.32	0.23
Percentage initially wrong, but subsequently correct						
Mean	45.57	31.34	32.60	39.99	28.87	33.24
s.e.	.38	.51	.57	.27	.42	.45
t	-22.37	-18.93	1.65	-22.27	-12.86	7.10
Regression residuals						
Mean	0.19	-0.74	-1.18	0.21	-1.14	-0.87
s.e.	.07	.12	.12	.03	.08	.07
t	-6.69	-9.86	-2.59	-15.80	-14.18	2.54

a/ The t-values placed in the "White" columns are for comparisons of the black and white students. The t-values in the "Black" columns are for comparisons of white and Hispanic students. The t-values in the "Hispanic" columns are for comparisons of blacks and Hispanics.

Sex

As shown in Table 5, females did not perform as well as males in either mathematics or science in either pre- or post-test administrations. For mathematics growth, the simple gains, percentage scoring higher on the post-test, percentage initially wrong but

subsequently correct, and regression residuals measures all agreed in showing less growth for females than for males. The science growth indicated nearly the same pattern (although the simple gain t-value fell just short of significance at the 0.05 level).

It should be noted that the elimination of ceiling effects caused the growth measure based on group mean differences to be identical with the measure based on simple gain scores. However, the use of between rather than within groups variance estimates effectively suppressed the t-values for the group mean difference measure.

Table 5.--Growth in computational mathematics and science, by sex

	Mathematics		Science	
Index	Male	Female	Male	Female
Base-year formula score				
Mean	11.06	10.23	9.82	8.58
s.e.	.16	.15	.08	.08
t	-3.78		-10.96	
First follow-up formula score				
Mean	12.83	11.59	10.83	9.48
s.e.	.19	.16	.08	.06
t	-4.99		-13.50	
Simple gain scores				
Mean	1.77	1.36	1.01	0.90
s.e.	.07	.07	.04	.04
t	-4.14		-1.94	
Group mean differences				
Mean	1.77	1.36	1.01	0.90
s.e.	.25	.22	.11	.10
t	-1.24		-.73	
Percentage scoring higher on the posttest				
Mean	64.46	61.24	57.98	56.19
s.e.	.64	.70	.50	.65
t	-3.39		-2.18	
Percentage initially wrong, but subsequently right				
Mean	44.61	39.97	39.50	36.26
s.e.	.54	.36	.35	.27
t	-7.15		-7.33	
Regression residuals				
Mean	0.18	-0.34	0.14	-0.30
s.e.	.08	.07	.04	.03
t	-4.89		-8.80	

SUMMARY

Scientists should be dismayed to observe that any pattern of growth may be found solely by judiciously selecting the "appropriate" growth measure. By examining the t-values in Tables 4 and 5, it was shown that some growth measures produced differences and others did not. Of course, some of the growth measures suffered from more problems than others.

Simple gain scores were plagued with the problem of reversal and attenuation. In many cases, the simple gain scores failed to identify differences which other measures yielded at high levels of significance. The group mean difference measure was even less powerful. The consistently lower t-values (for group mean differences relative to simple gains) are attributable to the use of between rather than within groups variance estimates.

In defense of the group means measure I should note that it is frequently used when the data are quasi-longitudinal. That is, for comparing pre- and post-tests when the group (or class) has experienced changes in composition. Dropouts and transfers (both in and out) can force analysts to either eliminate students from the comparisons or choose a measure like group mean differences. Its poor showing under the relatively ideal conditions described in this paper make me wonder how many true differences have been missed by this statistically powerless measure.

Percentage scoring higher on the posttest was an inconsistent measure. Reversal was probably a problem. Attenuation was also a probable problem after considering the expected values. Clearly, the expected value of 50 percent can be misleading to users.

Furthermore, the lack of any indication of the magnitude of the growth may be problematic.

The percentage of initially wrong items subsequently answered correctly measure did not appear to have reversal or attenuation problems. In addition, this measure consistently produced differences in the theoretically expected direction. The only problem with this measure is its computational complexity. The identification of the missed items on the pre- and post-tests is computationally more complex than the procedures used for the other measures. Frequently, data files for item responses are distinct from the general use files and this may cause hassles with merging and calculating this measure.

Regression residuals appeared to work well. Reversal was not a problem. Attenuation may have been a problem, but this measure consistently produced differences in the theoretically expected directions. The major problem with this measure is the form of the regression and agreement as to its specification. For this paper, simple weighted regression of the pre- test on the posttest was employed--but this is only one of an infinite family of regressions that could have been employed.

In conclusion, if you must study growth, be sure to use multiple measures. In fact, use as many measures as your budget will allow. When you can afford to use only one measure of growth, try to use either the percentage initially wrong, but subsequently correct measure or a regression residual, and hope!